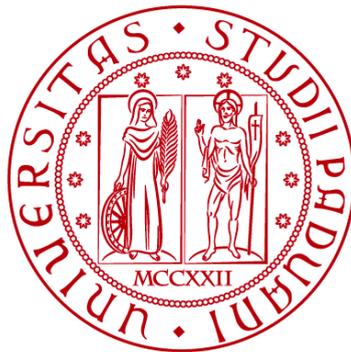


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea in Biologia**



**ELABORATO DI LAUREA**

**USO DI MODELLI METABOLICI PER LA  
FORMULAZIONE DI SINTROFIE NELLA  
DIGESTIONE ANAEROBICA**

**Tutor: Dott.ssa Laura Treu**

**Dipartimento di Biologia**

**Co-tutor: Dott. Nicola De Bernardini**

**Dipartimento di Biologia**

**Laureando: Vittorio Rossini**

**ANNO ACCADEMICO 2023/2024**

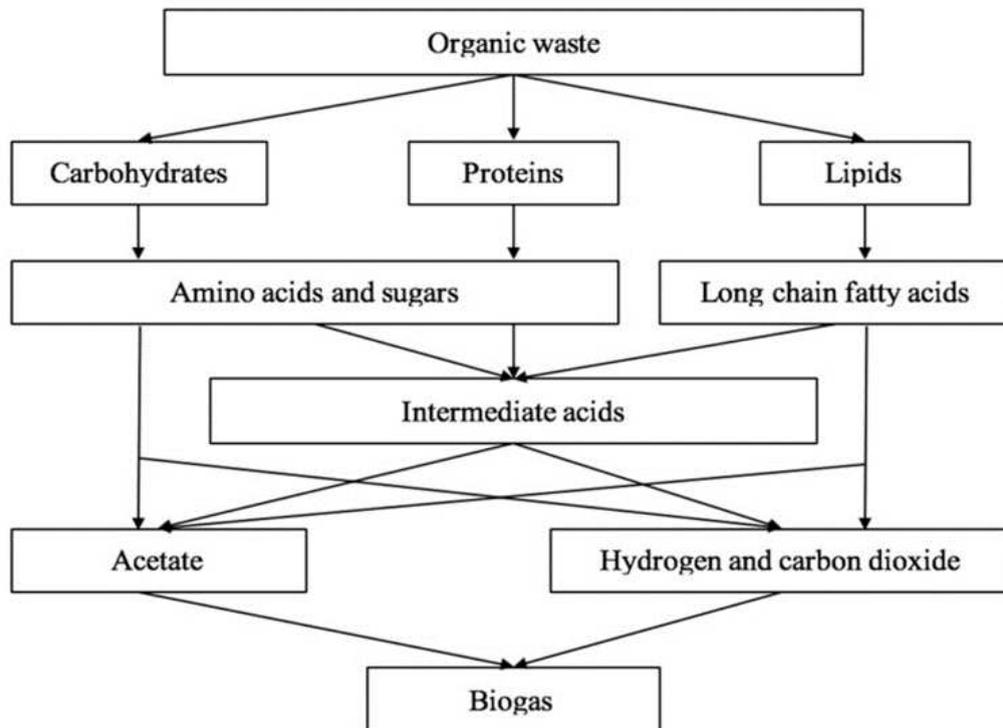


<b>1. Introduzione</b>	<b>4</b>
1.1 Digestione anaerobica e la sua rilevanza ambientale e industriale.	4
1.2 Definizione e importanza delle simbiosi nelle comunità microbiche anaerobiche	4
1.3 Complicazioni nel definire le simbiosi e utilità di approcci computazionali per la loro indagine	5
<b>2. Fondamenti dei Modelli Metabolici</b>	<b>7</b>
2.1 Introduzione a tecniche di modellazione metabolica	7
2.2 Revisione dello sviluppo di tecniche di modellazione e ruolo dell'integrazione di omiche	9
2.3 Benefici e sfide dell'integrazione di dati omici nei modelli metabolici	12
<b>3. Casi di studio nella modellazione simbiotica</b>	<b>13</b>
3.1 "Metabolic modeling of a mutualistic microbial community" di Sergey Stoliar et al. 2007.	14
3.2 "Integrating metagenomic binning with flux balance analysis to unravel syntrophies in anaerobic CO <sub>2</sub> methanation" di De Bernardini et al. 2023.	16
3.3 "Diverse electron carriers drive syntrophic interactions in an enriched anaerobic acetate-oxidizing consortium" di McDaniel et. al 2023	17
<b>4. Limiti e sviluppi futuri dei Modelli Metabolici</b>	<b>18</b>
4.1 Analisi critica dei limiti dei modelli metabolici attuali nella previsione del comportamento microbico reale	18
4.2 Discussione su possibili avanzamenti nella modellazione metabolica, inclusa l'integrazione di dati biologici più dinamici e approcci di apprendimento automatico.	21
4.3 Come questi miglioramenti potrebbero portare a una migliore accuratezza nella previsione delle interazioni microbiche e dei comportamenti dei sistemi	22
<b>5. Conclusione</b>	<b>22</b>
5.1 Riassunto delle principali scoperte della tesi	22
5.2 Riflessione sull'importanza dei modelli metabolici nel progredire la nostra comprensione delle interazioni simbiotiche e le loro implicazioni più ampie nella biotecnologia.	24
5.3 Utilizzo ed applicazione dei metanogeni	25

# 1. Introduzione

## 1.1 Digestione anaerobica e la sua rilevanza ambientale e industriale.

Affrontare le attuali sfide ambientali richiede un impegno deciso e coordinato. Si stima che entro il 2050 la popolazione mondiale raggiungerà i 9,6 miliardi, di cui il 66% vivrà in aree urbane. Attualmente in queste zone vive il 54% della popolazione e le loro economie rappresentano il 70% delle emissioni totali di CO<sub>2</sub> (Mangi, M.Y. et al.2020). Le attività umane concentrate in aree ristrette hanno spesso un impatto negativo sull'ambiente. Per mitigare questi effetti, è cruciale adottare un approccio proattivo che armonizzi le normative internazionali con l'azione delle comunità locali. Una soluzione flessibile consiste nell'implementazione di sistemi per la fissazione del carbonio mediante metanazione biologica. Possiamo inquadrare la DA come un processo biologico, oltre che un passaggio cruciale nel ciclo del carbonio, che converte la materia organica complessa in molecole più semplici in condizioni anossiche. Questo procedimento è naturalmente presente in molte nicchie ecologiche diverse (ad esempio, sedimenti marini, paludi terrestri, sistema digestivo degli eucarioti), pur essendo utilizzato intenzionalmente anche da approcci biotecnologici. Innanzitutto, la DA industriale viene sfruttata per generare biogas, un'alternativa rinnovabile al gas naturale composto da una miscela di metano (CH<sub>4</sub>) (50–75%) e anidride carbonica (CO<sub>2</sub>) (25–50%), tuttavia con quantità maggiore di contaminanti. Poiché la concentrazione di CH<sub>4</sub> generato è ottenuto dai reattori di biogas, raramente superiore al 60%, negli ultimi anni sono stati compiuti maggiori sforzi di ricerca per scoprire nuovi approcci in modo da accoppiare i processi di fissazione della CO<sub>2</sub> contenuta nel biogas, con la produzione di idrogeno (H<sub>2</sub>) verde. In questo modo il contenuto di CH<sub>4</sub> del biogas aumenta fino a valori di purezza prossimi al 100% e allo stesso tempo si ottengono tassi di produttività più elevati (De Bernardini et al).



**Figura 1.** Schema raffigurante in modo essenziale gli step necessari per la produzione finale di Biogas completamente riutilizzabile, dalla materia organica come punto di partenza.

Grazie alla sua duttilità nel recuperare nutrienti e carbonio per la fertilizzazione del suolo, di valorizzare i rifiuti organici e di produrre energia sotto forma di biogas, il processo deve essere accompagnato da una corretta programmazione. La teoria alla base della digestione anaerobica è stata definita da decenni; per questo tuttavia, gran parte della ricerca attuale è diretta all'ottimizzazione in diverse condizioni di digestione. Precisamente, questa tecnologia è stata implementata con successo nel trattamento dei rifiuti agricoli, degli scarti alimentari e dei fanghi delle acque reflue. La bioconversione è svolta da comunità microbiche complesse, che con un progressivo processo di degradazione degli scarti organici li trasformano in sostanze inorganiche come  $\text{CO}_2$  e  $\text{CH}_4$ . L'attività in situ dei singoli membri nelle comunità microbiche e le relazioni ecologiche esistenti tra i microbi, sono tuttora estremamente difficili da chiarire durante la digestione di substrati complessi (Zhu et al. 2020). Il nucleo di questa tesi vuole indagare i processi di simbiosi in comunità microbiche anaerobiche in grado di svolgere metanogenesi. Lo studio delle simbiosi è volto a

mettere a fuoco il potenziale metabolico dei singoli microrganismi, e contemporaneamente fornire nozioni. Specialmente perciò possiamo trasferire queste informazioni agli interi microbiomi, per analizzare complessivamente come funzionano le vie metaboliche della digestione senza ossigeno e le interazioni tra diversi microbi. Per ambire a ciò, proveremo a fare un passo indietro e concentrarci sul processo di DA e delle comunità microbiche che agiscono all'interno di essa. In aggiunta, al fine di sfruttare queste nozioni teoriche, andremo ad analizzare come l'applicazione di metodi: quali la metagenomica e nuovi metodi di analisi dei flussi, possano chiarire le potenzialità di specifiche comunità microbiche. Inoltre, la tesi chiarirà come l'avvento di nuovi metodi bioinformatici, ci abbia condotto alla possibilità di scoprire nuovi lignaggi, percorsi metabolici e interazioni microbiche. Essa avviene attraverso quattro fasi successive: idrolisi, acidogenesi, acetogenesi e metanogenesi. Essenzialmente diversi sono i primi tre step dall'ultimo, in quanto essi sono svolti da gruppi batterici, mentre la quarta fase, da archaea. I digestori anaerobici tipicamente incontrano biomassa organica, che contiene polimeri complessi inaccessibili ai microrganismi senza essere ulteriormente scomposti attraverso l'idrolisi o pretrattamenti. Di conseguenza, il processo iniziale di idrolisi ha lo scopo di scindere macromolecole organiche nei suoi componenti più piccoli, che a loro volta possono essere utilizzati dai batteri acidogeni. I batteri idrolitici hanno la peculiarità di secernere enzimi extracellulari che possono convertire carboidrati, lipidi e proteine in zuccheri, acidi grassi a catena lunga (LCFA, long chain fatty acid) e aminoacidi, rispettivamente. È importante notare che alcuni substrati, come la lignina, la cellulosa e l'emicellulosa, sono difficilmente degradabili perché le loro strutture rendono in gran parte inaccessibili punti di legame per gli enzimi (Lin et al. 2010). Assorbendo i prodotti dell'idrolisi attraverso le loro membrane cellulari, i microrganismi acidogeni sono in grado di produrre acidi grassi volatili (VFA, volatile fatty acid). I VFA costituiscono una classe di acidi organici come l'acetato, il propionato e il butirato, essenziali per l'equilibrio del processo. A differenza degli altri stadi, si ritiene generalmente che il secondo step, ovvero la acidogenesi, proceda a un ritmo più rapido di tutte le altre fasi della digestione anaerobica, con i batteri acidogeni che hanno un tempo di duplicazione inferiore alle 36 ore. Tenendo presente la rapidità di questa fase, è importante notare che durante la produzione dei VFA si assiste alla creazione di precursori diretti della metanogenesi. Oltretutto l'acidificazione dovuta da eccessivo accumulo di VFA è segnalata come causa frequente di inibizione degli archaea. A questo punto è inoltre opportuno esaminare il processo di degradazione degli aminoacidi, in quanto la loro degradazione può portare a conseguenze opposte. Il diretto prodotto di scarto della deaminazione è la produzione di ammoniaca, che a

concentrazioni sufficientemente elevate, aumenta il pH e funge da inibitore della DA. Infine, in questa fase è nondimeno cruciale la concentrazione dell'idrogeno, un prodotto intermedio di molte reazioni, in quanto esso controlla l'efficienza termodinamica di degradazione di aminoacidi e VFA. Con la produzione di acetato attraverso l'acidogenesi, si perde una porzione del substrato originale che è già stato trasformato in un substrato adatto alla metanogenesi acetoclastica (Fournier e Gogarten 2008). Tuttavia, i VFA prodotti in quantità più elevate devono ancora essere resi accessibili ai microrganismi metanogeni. L'acetogenesi pertanto è il processo mediante il quale questi VFA più elevati e altri intermedi vengono convertiti in acetato, con cui viene prodotto anche l'idrogeno. L'idrogeno prodotto durante l'acetogenesi apre la discussione su un argomento cruciale. Mentre questa fase è produttrice di idrogeno, una pressione parziale eccessiva si rivela deleteria per i microrganismi acetogeni (Dinopoulou G et al. 1988), infatti è importante che la collaborazione tra gli organismi che effettuano le reazioni di ossidazione anaerobica e il gruppo di quelli che formano metano, dipenda dalla pressione parziale dell'idrogeno presente nel sistema. Tuttavia, a causa della presenza di metanogeni idrogenotrofi, l'idrogeno può essere consumato rapidamente mantenendo le pressioni parziali a un livello favorevole all'acetogenesi creando una reazione esoergonica (Meegoda et al. 2018). La metanogenesi segna la fase finale della digestione anaerobica, in cui gli intermedi accessibili vengono consumati dai microrganismi metanogeni per produrre metano (James G. Ferry, 2010). I microrganismi metanogeni rappresentano un gruppo di archaea anaerobi obbligati e, oltre alla sensibilità all'ossigeno, essi hanno la peculiarità di essere confinati ad una ristretta cerchia di substrati che ne limita la capacità operativa. Per quanto riguarda le esigenze ambientali della metanogenesi, questi microorganismi tendono a richiedere un pH più elevato rispetto alle fasi precedenti della digestione anaerobica, oltre a un potenziale redox inferiore, requisito quest'ultimo che ha causato notevoli problemi alla coltivazione in laboratorio (Wolfe RS. 2011). Allo stesso tempo però, i metanogeni sembrano avere un tempo di rigenerazione significativamente più lento rispetto ad altri microrganismi nella digestione anaerobica, fino a 5-16 giorni.

## 1.2 Definizione e importanza delle sintrofie nelle comunità microbiche anaerobiche

Secondo la legge dell'equilibrio chimico, il consumo dei metaboliti escreti da parte dei microbi riceventi può accelerare il metabolismo dei microbi donatori. Questo è il concetto di sintrofia, che è un tipo di mutualismo, uno dei più comuni all'interno della biosfera, in cui due o più specie diverse che vivono in stretta vicinanza fanno affidamento l'uno sull'altro per nutrienti, protezione e/o altre

funzioni vitali (Boucher D.H. 1985). Nonché cruciale nel governare il metabolismo e la crescita di diversi microbi negli ecosistemi naturali e ingegnerizzati . Un esempio nelle comunità metanogeniche, la riduzione degli equivalenti, come idrogeno e formiato, è mediato dal trasferimento tra partner sintrofici. Gli studi hanno rivelato che i microbi coinvolti nella sintrofia abbiano evoluto meccanismi molecolari per stabilire partnership specifiche e comunicazioni interspecie, con una conseguente operazione metabolica efficiente. E bene precisare che una ben definita sintrofia si viene a stabilire tra batteri fermentativi (sintrofi) e metanogeni archaea (metanogeni), che cooperativamente trasformano il materiale organico composto quali acidi grassi volatili (VFA, compreso butirrato, propionato, e acetato) in metano (Schink, 1997). Approfondendo l'interazione sintrofica tra questi microbi, essa si basa sull'equivalente di riduzione del trasferimento di sostanze, che è anche chiamato "trasferimento elettronico interspecie (IET, Interspecies electronic transfer)". La relazione dal punto di vista chimico, deriva dalla bassa resa energetica della degradazione metanogenica rispetto a quella dei processi ossidativi alternativi che perciò può essere la ragione per cui la metanogenesi è l'ultima fase che si verifica, dopo che gli altri accettori di elettroni sono stati ridotti. Come conseguenza di questo piccolo guadagno energetico, il prodotto di reazione, il metano, immagazzina una parte importante dell'energia disponibile nella conversione della biomassa aerobica. Questo guadagno può essere successivamente sfruttato per la produzione di energia in presenza di ossigeno da parte di altri organismi, ad esempio, da ossidanti aerobici del metano o da parte dell'uomo per il riscaldamento o altri processi fisici. La piccola quantità di energia disponibile nella conversione metanogenica costringe i microrganismi coinvolti in una cooperazione molto efficiente. La dipendenza reciproca dei batteri partner rispetto alla limitazione di energia può essere così forte che nessuno dei due partner può operare senza l'altro e che insieme, esibiscono un'attività metabolica che presi singolarmente non potrebbero realizzare. Caso specifico è la crescita dei metanogeni idrogenotrofi, dipendente dalla fornitura di H<sub>2</sub> da parte dei sintetizzatori che producono idrogeno, questi due gruppi di organismi hanno un legame nutrizionale reciproco (Jackson e McInerney, 2002). Tali cooperazioni sono chiamate relazioni sintrofiche (Schink 1997).

### 1.3 Complicazioni nel definire le sintrofie e utilità di approcci computazionali per la loro indagine

I progressi recenti nell'uso del sequenziamento ad alto rendimento e di tecniche

di analisi di tutta la comunità come la meta-genomica e la meta-trascrittomica promettono di rivoluzionare la disponibilità di informazioni genomiche. Nonostante ciò però sappiamo ancora molto poco circa i contributi metabolici dei singoli operatori microbici all'interno di una nicchia ecologica, tra cui l'estensione e la direzionalità delle interazioni tra di loro. Si può notare dunque la necessità di una crescita nello sviluppo di strutture di modellazione efficienti per far luce sugli aspetti meno compresi del metabolismo nelle comunità microbiche. Malgrado questi sforzi, tutti i metodi esistenti per l'analisi del bilancio di flusso delle comunità microbiche si basano su problemi di ottimizzazione con una singola funzione oggettiva (relativa alle singole specie), che non può sempre catturare la natura multi-livello della decisione-produzione in comunità microbiche. Ad esempio, il modello di analisi del bilancio di flusso descritto in uno dei casi di studio che andremo ad affrontare, è applicabile solo alle associazioni sintrofiche, in cui la crescita di entrambe le specie è accoppiata attraverso il trasferimento di un metabolita chiave. Per far fronte a ciò, molti modelli computazionali sono stati sviluppati ed una soluzione innovativa e attuale può essere un software come OptCom, un framework completo per l'analisi della bilancia di flusso per le comunità microbiche, che si basa su una descrizione di ottimizzazione multi-livello. Contrariamente agli approcci precedenti che si basano su una singola funzione oggettiva, la struttura multi-livello/obiettivo di OptCom consente di valutare correttamente i compromessi tra i criteri di idoneità individuali vs. comunitari. Questa struttura di modellazione è abbastanza generale da catturare qualsiasi tipo di interazione (positiva, negativa o combinazione di entrambe) per qualsiasi numero di specie coinvolte. Inoltre OptCom, è in grado di spiegare in vivo le osservazioni di livelli di ottimalità della crescita per ogni partecipante della comunità. In particolare, si presta come un tool di nuova generazione per la sua capacità nell'essere utilizzato per valutare il livello ottimale di crescita per diversi membri in una comunità microbica (cioè, modalità descrittiva) e successivamente fare previsioni per quanto riguarda il traffico metabolico (cioè, modalità predittiva) dati i livelli di ottimalità identificati (Zomorodi e Maranas 2012). Attualmente sono disponibili diversi strumenti di ricostruzione metabolica, ognuno dei quali offre diversi gradi di compromesso tra automazione e intervento umano. Questi strumenti seguono un approccio di ricostruzione bottom-up costituito dai seguenti passaggi principali: annotare geni con funzioni metaboliche; recuperare le rispettive reazioni biochimiche da un database di reazione, come KEGG (Kanehisa M et al. 2000); assemblare un progetto di rete metabolica ed infine curare manualmente il modello di bozza. L'ultimo passo include diversi compiti, come l'aggiunta di reazioni mancanti necessarie per generare precursori della biomassa (riempimento del gap), la

correzione dell'equilibrio elementare e della direzionalità delle reazioni, la rilevazione di cicli futili e la rimozione di reazioni bloccate e metaboliti morti. Se questi problemi non vengono risolti, il modello può generare previsioni non realistiche del fenotipo, come rendimenti errati della biomassa, generazione eccessiva di ATP, falsa essenzialità genica o requisiti nutrizionali errati. La fase di cura manuale richiede molto tempo e include compiti ripetitivi che devono essere eseguiti per ogni nuova ricostruzione. A questo problema c'è un ulteriore software inquadrato come possibile soluzione, CarveMe, un nuovo strumento di ricostruzione che implementando un approccio di ricostruzione top-down, automatizza la creazione di modelli di comunità microbiche fondendo gruppi selezionati di modelli di singole specie in reti su scala comunitaria. (Machado D, et al. 2018) Software sono stati sviluppati anche per far fronte alla problematica dell'immensità di dati da elaborare previo lo studio dei mutualismi nelle comunità batteriche. Per esempio senza considerare i metodi di sequenziamento per generare dati, le prime fasi nell'analisi dell'interezza di un genoma comprendono il confronto delle sequenze da analizzare, con quelle già conosciute e presenti nel database. Questo compito computazionalmente intensivo fornisce i tipi di dati di base per molte analisi successive, tra cui confronti filogenetici, annotazioni funzionali, binning di sequenze, profilazione filogenetica e ricostruzioni metaboliche. Viene proposto un sistema open source liberamente disponibile e completamente automatizzato per l'elaborazione dei dati della sequenza del metagenoma per generare questi elementi di base. Un'implementazione pubblica di questo sistema è stata fornita affinché tutti i ricercatori analizzino in autonomia i loro metagenomi. Il servizio, il server RAST di metagenomica (mg-RAST in breve), è disponibile sul web a tutti i ricercatori e l'accesso non è limitato a gruppi specifici o tipologie di dati. Per attestare la bontà del software, finora sono stati elaborati quasi 500 metagenomi attraverso la versione beta della pipeline (Meyer et al. 2008).

## 2. Fondamenti dei Modelli Metabolici

### 2.1 Introduzione a tecniche di modellazione metabolica

Il numero e la portata dei metodi sviluppati per interrogare e utilizzare le ricostruzioni delle reti metaboliche si è notevolmente espansa, come le aree in cui queste hanno trovato terreno fertile per il loro utilizzo. In particolare, la ricostruzione delle reti è stata utilizzata per affrontare un ampio spettro delle applicazioni di base e pratiche in cinque categorie principali: ingegneria metabolica, scoperta modello-diretta, interpretazioni di schermi fenotipici, analisi di proprietà di rete e studi di processi evolutivi. Stimolato da queste realizzazioni,

si prevede di proseguire ed ampliare ulteriormente la portata e il contenuto delle ricostruzioni di reti, sviluppare nuove strumentazioni di analisi in silico, ed espandersi in adattamento agli usi della causalità prossimale e distale in biologia. Presi insieme, questi sforzi rafforzeranno il rapporto meccanicistico tra genotipo-fenotipo per la definizione e lo studio del metabolismo microbico (Orth et al. 2010). I modelli stechiometrici globali del metabolismo microbico sono riusciti a prevedere tali proprietà emergenti, in particolare per gli organismi ben studiati (Price et al 2004). L'ingrediente principale di un modello stechiometrico è la completa descrizione della connettività biochimica tra i metaboliti cellulari (detti matrice stechiometrica), che è paragonabile alle classiche mappe biochimico-pathway. Prima che una ricostruzione possa essere usata per calcoli di rete e/o capacità fisiologiche, c'è un passo all'apparenza irrilevante, ma essenziale, in cui una ricostruzione viene convertita in una rappresentazione matematica. Questa conversione traduce un "GENERE" (contenitore di tutte le trasformazioni chimiche che avvengono nella rete) in un formato matematico che diventa la base per un modello di scala genomica (GEM). Queste trasformazioni possono essere rappresentate stechiometricamente e successivamente formano una matrice, le cui file rappresentano i composti, mentre le colonne rappresentano le trasformazioni chimiche e le entrate sono i coefficienti stechiometrici (Price et. al 2010). Questa consente il possibile utilizzo di una vasta gamma di strumenti computazionali per analizzare le proprietà della rete. Essi si concentrano sulla valutazione delle proprietà sistemiche e sulle funzioni che può svolgere sotto i vincoli fisico-chimici posti sulla cellula. A questo step compete la cosiddetta ricostruzione basata sul vincolo e l'analisi della struttura (Reed e Palsson 2003) per l'organismo bersaglio (COBRA Toolbox). Al termine di questa fase, può essere applicato un modello per studiare la condizione di crescita specifica da cui sono stati basati i dati di formazione e può essere sfruttato per esplorare condizioni ambientali aggiuntive. I set di dati ad alto rendimento, High-throughput data sets, valutano un gran numero di interazioni attraverso diverse condizioni di crescita o genetiche, in aggiunta possono essere utilizzati per perfezionare ed estendere il contenuto metabolico di una rete. Questo tipo di confronti ed analisi hanno il potenziale di valutare la totalità dei dati omici della scala genomica in un modo integrato disponendoli in un contesto funzionale e strutturato. Per arrivare alle diverse scoperte, viene analizzata una vasta gamma di dati (ad esempio, fenotipizzazione, espressione genica e attività enzimatica) per ipotizzare e convalidare le previsioni computazionali. Questo approccio dimostra come la premessa di integrare risultati di modellazione e dati sperimentali, diventerà un approccio chiave per espandere le attuali conoscenze metaboliche insieme alla scoperta di nuovi componenti e interazioni nei processi

cellulari. Anche se la genomica genetica e la metabolomica, analisi della concentrazione e della dinamica di piccole molecole cellulari chiamate metabolomi, forniscono un inventario dei metaboliti e suggeriscono le relazioni chimiche possibili fra loro, le informazioni dinamiche dall'analisi di flusso possono essere usate per perfezionare ulteriormente i modelli stechiometrici della reazione. La metodologia principe per ambire a ciò è la FBA, flux balance analysis, un approccio largamente utilizzato per lo studio e la comprensione delle dinamiche biochimiche che intercorrono tra le specie batteriche e non. Queste ricostruzioni contengono tutte le reazioni metaboliche note in un organismo ed i geni che codificano ogni enzima. FBA, in particolare, calcola il flusso dei metaboliti attraverso questa rete metabolica, rendendo così possibile prevedere il tasso di crescita di un organismo o la produzione di una biotecnologia metabolita importante. Lo strumento principale su cui si basa questa rappresentazione è una tabulazione, sotto forma di matrice numerica, dei coefficienti stechiometrici di ciascuna reazione. I vincoli come questi si trovano al cuore della FBA, differenziando l'approccio da modelli basati sulla teoria dipendenti da equazioni biofisiche che richiedono molte difficoltà nel misurare i parametri cinetici. Tuttavia la FBA possiede delle limitazioni. Poiché non usa parametri cinetici, non può predire le concentrazioni del metabolita ed infatti è adattato in gran parte alla determinazione dei flussi allo stato stazionario. Tranne che in alcune forme modificate, la FBA non tiene conto degli effetti regolatori come l'attivazione degli enzimi da parte delle chinasi proteiche o la regolazione dell'espressione genica, quindi le sue previsioni potrebbero non essere sempre accurate (Orth et al. 2010).

## 2.2 Revisione dello sviluppo di tecniche di modellazione e ruolo dell'integrazione di omiche

L'integrazione di omiche, mira a combinare due o più insiemi di set di dati omici per aiutare nell'analisi, nella visualizzazione e nell'interpretazione di informazioni volte a determinare il meccanismo di un trattamento biologico. Gli obiettivi dell'integrazione delle basi di dati consistono nel rendere i dati disponibili in modo più completo e nell'aumentare la qualità dei dati esistenti, consentendo di esaminare questioni precedentemente difficili da affrontare. Tuttavia, il proliferare di una miriade di strumenti, set di dati e nuovi approcci tende ad inondare la letteratura e sopraffare i nuovi ricercatori del campo (Krassowski et al. 2020). Integrare grandi quantità di dati eterogenei costituisce non solo una

sfida concettuale, ma un ostacolo pratico nell'analisi quotidiana dei dati omici. Con l'avvento di nuove tecnologie omiche e attraverso progetti di grandi consorzi, i sistemi biologici vengono ulteriormente studiati su una scala senza precedenti generando insiemi di dati eterogenei e spesso di enormi dimensioni. Questi set di dati automaticamente, incoraggiano i ricercatori a sviluppare nuove metodologie di integrazione. (Fournier e Gogarten 2008) Nell'ultimo decennio, l'applicazione di diversi studi omici presi individualmente (ad esempio, genomica, epigenomica, trascrittomica, proteomica, metagenomica) che miravano a comprendere un particolare problema nella malattia umana (Karczewski e Snyder, 2018), agricoltura (Ichihashi et al., 2020), scienza delle piante (Liu et al., 2016), microbiologia (Quinn et al., 2016), e l'ambiente hanno avuto un grande successo. Ciascuna di queste tecniche è progredita rapidamente e spingono queste scienze ad un continuo progresso:

### 1) Genomica

È la più longeva tra tutte le tecnologie omiche e si riferisce allo studio sulle sequenze intere del genoma e sulle varianti di sequenza del DNA in esso, comprese le singole variazioni del nucleotide, l'inserzione-cancellazioni, le variazioni strutturali e le alterazioni del numero della copia. Con l'avvento delle tecnologie di sequenziamento di prossima generazione (NGS) negli ultimi due decenni, i genomi possono ora essere analizzati più velocemente, in modo più economico e in modo ad alta produttività. I costi di ordinamento del genoma sono precipitati costantemente dai miliardi di dollari per ordinare il genoma umano iniziale in 2000 a appena \$100 per il genoma in 2022 (Ultima Genomics).

### 2) Epigenomica

Con l'ausilio di molteplici multiple, i modelli mancanti o con una deficienza nei dati di espressione genica possono essere compensati, per esempio, dall'epigenomica (Zarayeneh et al. 2017). Si riferisce alla catalogazione completa delle modifiche chimiche del DNA e degli istoni a cui gravita intorno. L'ambiente dell'epigenomica ha cominciato a prendere piede con la scoperta della metilazione del DNA e delle modifiche dell'istone negli anni 60, ed è stato accelerato dalle tecnologie di NGS. Diverse tecniche NGS come la metilazione del DNA hanno permesso una mappatura precisa dei modelli di metilazione del genoma e di altri marcatori epigenetici che influenzano la regolazione genica.

### 3) Proteomica

Definita come la quantificazione di tutta l'identità e l'abbondanza della proteina in un campione, ha visto i propri progressi più consistenti nelle tecnologie e nella strumentazione, permettendo una rilevazione più veloce, più efficiente, sensibile ed accurata delle proteine. La prima analisi ad alta definizione della lavorazione di proteine, è stata raggiunta facendo uso delle metodologie "protein array", basate su chip prefabbricati con rilevazione specificata della proteina. Nonostante la sensibilità del processo, questo approccio vide difficoltà nel catturare l'intero proteoma. Il ventunesimo secolo tuttavia, ha testimoniato i miglioramenti significativi sia nella cromatografia liquida (LC) che nei parametri della spettrometria di massa (MS), particolarmente nella frequenza di scansione più alta e nell'accuratezza di massa, promettendo una futura generazione proteomica basata sugli approcci LC-MS/MS.

### 4) Trascrittomica

Scienza che fornisce dati riguardo alla precisa espressione di mRNA delle cellule. Data la ricchezza di dati trascrittomici, gli sforzi per integrare i dati di espressione mRNA con ricostruzioni di rete metaboliche, in particolare, hanno fatto progressi significativi quando si utilizza come piattaforma analitica una FBA. A sostegno di ciò, negli ultimi anni si è assistito a parecchi avanzamenti nell'integrazione dei dati transcriptomici con le ricostruzioni delle reti metaboliche. In particolare, sono stati introdotti numerosi algoritmi basati su FBA che utilizzano i livelli di trascrizione mRNA derivati sperimentalmente per modificare le reazioni della rete inattivandole interamente o limitando i loro livelli di attività.

### 5) Metabolomica

Essa si riferisce allo studio di piccole molecole ed ha osservato un miglioramento esponenziale nelle tecnologie e nella strumentazione negli ultimi decenni. Gli approcci importanti di metabolomics comprendono il metabolomics mirato, il metabolomics untargeted, il fluxomics e la rappresentazione del metabolita. Il metabolomics mirato mira ad identificare e quantificare un piccolo sottoinsieme dei metaboliti (50-500) ed è ideale per rilevazione di biomarcatore. Il metabolomics untargeted tenta di caratterizzare tutto il numero possibile di metaboliti (>10.000). Fluxomics invece è un ramo del metabolomics mirato che controlla il movimento dei contrassegni isotopici con i mediatori metabolici e misura i tassi di reazione del metabolita. La rappresentazione del metabolita è un campo emergente di metabolomics che comprende la rilevazione e la visualizzazione dei metaboliti in tessuti. Questi studi generano una pleora di dati,

che, con un'attenta integrazione in un adeguato quadro statistico e matematico, possono aiutare a risolvere più ampie questioni relative alle aree di base e applicate della biologia. Per di più, da un lato supportano i ricercatori nella riproduzione e convalida dell'analisi di altri laboratori, e d'altra parte permettono a quest'ultimi di analizzare i dati in modi nuovi e/o con metodologie diverse che non erano originariamente considerati dal team che ha generato le informazioni specifiche.

### 2.3 Benefici e sfide dell'integrazione di dati omici nei modelli metabolici

La complessità dei dati a cui facciamo riferimento sono database di tipo "eterogeneo", in cui sono coinvolte due o più fonti di dati fondamentalmente diverse. Ciò comporta la conversione di ciascuna delle fonti separate in struttura, formato e dimensione comuni prima di combinarle. La normalizzazione dei dati omici comprende l'utilizzo di vari strumenti, con ciascuno adattato a tipi di dati specifici ed ai requisiti analitici. La normalizzazione basata a tendenza centrale, quale la media e la mediana, è un metodo semplice ma efficace impiegato per la normalizzazione dei dati di metabolomici e di proteomici per esempio. Riscrive i valori di intensità dei singoli campioni per allinearli con l'intensità media o mediana su tutti i campioni. Ciascuno di questi tipi di dati distinti fornisce un punto di vista differente, parzialmente indipendente e complementare dell'intero genoma. Tuttavia, le funzioni di comprensione dei geni, delle proteine e di altre funzioni del genoma richiedono più informazioni di quelle fornite da ciascuno degli insiemi di dati e devono inoltre superare diverse complicazioni lungo il loro percorso (Hamid et al. 2009). I dati genomici si presentano sotto forma di vettori, grafici, o sequenze quindi, è di cruciale importanza considerare con attenzione le strategie che catturano le informazioni contenute in ogni tipologia di dato, prima di combinarli tra loro. Occuparsi dei valori mancanti è una funzione critica dei dati di omici, che se non aggiustati adeguatamente, possono insidiare l'esattezza e l'affidabilità dei risultati. Per attenuare questo problema, i metodi di imputazione sono comunemente impiegati per stimare i valori mancanti e minimizzare l'impatto della scarsità di dati. (Flores et al. 2023) Prendendo in esame la modellazione metabolica su scala genomica (GSMM), essa si presenta come una tecnica di modellazione matematica basata sul vincolo che è stata determinante nell'analisi dei dati omici (Palsson et al. 2010)(Brunk et al. 2018). I modelli metabolici a scala genomica (GEMS) forniscono una struttura robusta che permette l'integrazione di insiemi di dati omici multipli. Sfruttando la potenza dei GEMS, i ricercatori possono approfondire le complessità dei percorsi biologici, consentendo una comprensione completa del metabolismo cellulare e dei suoi meccanismi sottostanti. I GEMS sono stati ampiamente utilizzati nell'ingegneria

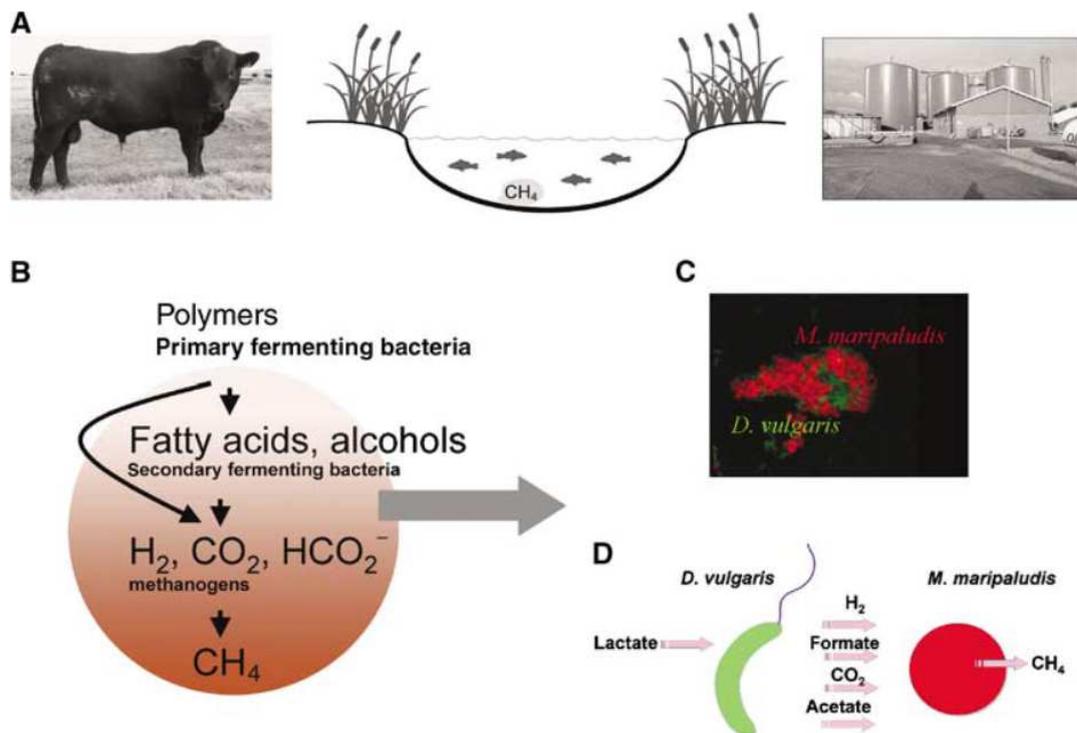
metabolica, dove hanno dimostrato la loro capacità di prevedere la crescita cellulare in diverse condizioni nutrizionali. Inoltre, sono stati applicati a studi di progettazione, studiando l'essenzialità delle reazioni/geni, la rilevanza delle vie metaboliche, modellando i fenotipi e manipolando questi pathway(s). Negli ultimi anni, vari gruppi hanno introdotto strutture basate sull'ottimizzazione per prevedere modifiche genetiche (cioè, deletions and additions), volte a massimizzare la secrezione di agenti biochimici dalle reti metaboliche. L'obiettivo era quello di guidare le strategie di ingegneria metabolica sperimentale adottando un approccio sistemico per anticipare l'effetto delle modifiche genetiche sul metabolismo. Il concetto che sta dietro a questo quadro computazionale si basa sul fatto che se una reazione è considerata up- o down regulated se è costretta ad assumere valori di flusso significativamente superiore o inferiore al suo stato stazionario prima delle manipolazioni genetiche. Mentre la disponibilità di dati genomici è ragionevolmente ben fornita da archivi di dati accessibili al pubblico e ben mantenuti (con la relativa eccezione dei dati clinici), vi è la necessità di migliorare gli standard di annotazione e i requisiti negli archivi di dati per consentire una migliore integrazione e riutilizzo dei dati disponibili pubblicamente. L'aspetto di sfruttamento dei dati dell'integrazione dei dati è probabilmente quello che richiede maggiore attenzione, in quanto coinvolge l'uso di conoscenze pregresse - e la sua conservazione efficiente, lo sviluppo di metodi statistici per analizzare insiemi di dati eterogenei e la creazione di strumenti di esplorazione dei dati che incorporano sia statistiche di riepilogo utili che nuovi strumenti di visualizzazione. In conclusione, incorporando nuovi set di dati disponibili e ben annotati nei GEMS esistenti è possibile affrontare efficacemente le sfide associate alla completezza del modello. Inoltre, l'incorporazione di dati multi-omici non solo facilita il perfezionamento e la valutazione di questi modelli, ma ne migliora anche la convalida, aumentando così la nostra fiducia nelle previsioni future.

### 3. Casi di studio nella modellazione sintrofica

L'analisi dei modelli metabolici è stata portata avanti per lo studio dei mutualismi tra le colture batteriche, è stata portata avanti attraverso la meta-analisi dei seguenti tre articoli, precisamente selezionati per l'ampia varietà degli approcci sfruttati. L'obiettivo comune degli studi in analisi è indagare i flussi di metaboliti tra specie microbiche di diverse colture coinvolte nel processo di ....

### 3.1 "Metabolic modeling of a mutualistic microbial community" di Sergey Stoloyar et al. 2007.

In questo studio gli autori esplorano il modello metabolico di una comunità microbica che include 2 specie: il batterio *Desulfovibrio vulgaris* e l'archea *Methanococcus maripaludis*. Lo scopo principale del lavoro è sviluppare un modello metabolico della comunità in grado di rappresentare in silico l'associazione sintrofica che si instaura tra le specie. Il modello ha inoltre lo scopo di prevedere le interazioni metaboliche e la crescita della co-coltura in condizioni specifiche. Il tutto è stato reso possibile dal previo sequenziamento dei due genomi, andando a prevedere diverse caratteristiche ecologicamente rilevanti. Il modello generato si basa sui modelli dei metabolismi centrali annotati dal sequenziamento dei genomi di entrambi gli organismi. Nel modello, il lattato è stato fornito come unica fonte di carbonio ed energia per la crescita della co-coltura oltre che per la crescita di *Desulfovibrio* in monocoltura su solfato. Mentre durante la crescita sintrofica, è possibile che *M.maripaludis* possa anche usare l'acetato prodotto da *D. vulgaris* come fonte di carbonio. L'approccio ha permesso di prevedere accuratamente l'abbondanza relativa di *D. vulgaris* rispetto a *M.maripaludis* durante la crescita (*D. vulgaris*, domina numericamente con un rapporto di 2:1 nella co-coltura), nonché i flussi di metaboliti scambiati tra loro. Una delle fondamentali scoperte è stata che lo scambio di idrogeno è essenziale per la crescita sintrofica, mentre il formato, escreto da *D.vulgaris* per l'utilizzo da parte di *M.maripaludis* durante la crescita in co-coltura potrebbe non essere necessario come veicolo interspecie di elettroni, nonostante sia teoricamente possibile. Infatti l'ipotesi principale su cui si sviluppa l'intero articolo è che in assenza di un accettore di elettroni adatto per il riduttore di solfato, i metanogeni creano favorevoli condizioni termodinamiche eliminando idrogeno e mantenendo bassa la pressione parziale di  $H_2$ , questo ambiente permette ai riduttori di solfato di utilizzare la fonte di carbonio in un processo simile alla fermentazione per crescere e svilupparsi.



**Figura 2. Interazioni vitali per la produzione di metano:** nei ruminanti, nei sedimenti d'acqua dolce e nei digestori anaerobici, le interazioni tra i microrganismi che producono idrogeno come *D. vulgaris* e gli archaea metanogenici favoriscono attivamente la metanogenesi. Questo processo sfrutta l'idrogeno e altri composti derivati dalla fermentazione per ridurre l'anidride carbonica a metano, mostrando come *D. vulgaris* trasferisca composti a *M. maripaludis* per la produzione di energia.

In particolare sono stati sfruttati modelli di bilancio di flusso (FBA, flux balance analysis) per diversi organismi in modo da definire la gamma di fenotipi teoricamente possibili per un dato genotipo simulando le condizioni ambientali o bloccando i flussi attraverso percorsi particolari (Edwards and Palsson, 2000; Klamt et al, 2002; Van Dien and Lidstrom, 2002; Borodina et al, 2005; Feist et al, 2006). Infatti è stato possibile raggiungere tale risultato in particolare separando i metaboliti in tre compartimenti, tra i quali non è necessariamente libero lo scambio di metaboliti. Infine il lavoro dimostra l'utilità dei modelli metabolici e dei modelli stechiometrici per prevedere i flussi metabolici e i fenotipi di crescita di organismi singoli, ma anche per catturare i parametri di crescita e le variazioni di struttura di comunità microbiche semplici

### 3.2 “Integrating metagenomic binning with flux balance analysis to unravel syntrophies in anaerobic CO<sub>2</sub> methanation” di De Bernardini et al. 2023.

In questo secondo studio si osserva una notevole evoluzione nel processo di ricostruzione dei modelli metabolici di comunità, con particolare attenzione alla loro funzione nel chiarire i meccanismi sintrofici coinvolti nella biometanazione della CO<sub>2</sub>. Lo studio ha selezionato come caso di studio il biofilm, generato durante il processo di upgrading del biogas. L'avanzamento metodologico qui osservato, deriva dall'integrazione di dati di sequenziamento all'avanguardia. Un ruolo fondamentale all'interno di questa analisi inoltre è stato svolto da tool automatizzati come Gap Seq e le sue estensioni, fornendo una panoramica completa delle capacità metaboliche dei microbiomi e permettono di esplorare nel profondo le interazioni e le funzioni ecologiche all'interno di comunità microbiche complesse. Mediante l'uso integrato del sequenziamento Nanopore e Illumina, lo studio ha ricostruito 59 genomi assemblati da metagenomi (MAGs), di cui però solo il 60% di alta qualità, che hanno permesso di delineare con precisione: le specie predominanti nel biofilm. Due in particolare: *M. wolfeii* GSMM957 e *Limnochordiasp.* GSMM975 hanno rivelato essere le più abbondanti e avere metabolismi molto complessi. È stato scoperto, attraverso l'analisi approfondita dei geni chiave per la produzione del substrato solido, che queste specie si vedono scambiare metaboliti essenziali come idrogeno, anidride carbonica e formiato, cruciali per la conversione efficiente del CO<sub>2</sub> in CH<sub>4</sub>, ottimizzando così la qualità del biogas prodotto. Un altro aspetto di particolare rilevanza è stato il ruolo di alcuni microrganismi (*Firmicutes sp.* GSMM966 in particolare) nella escrezione di sostanze polimeriche extracellulari (EPS), le quali hanno la capacità di generare materiale cruciale, per lo sviluppo di un ambiente di crescita protetto (Langer S, et al. 2014), come quello accumulato sulla membrana e sul diffusore di upgrading dell'esperimento proposto. Da sottolineare, come l'analisi del contributo di ciascuna specie alla sintesi dell'EPS, sia importante per far luce sui potenziali meccanismi che possono essere sfruttati per manipolare le comunità microbiche e influenzare i sistemi AD. L'applicazione dei modelli di bilancio dei flussi, per prevedere il comportamento microbico sotto l'applicazione di condizioni sperimentali, ha inoltre rivelato nuove comprensioni sulle interazioni microbiche, evidenziando l'importanza del trasferimento di formiato tra le specie, che facilita processi sintrofici vitali per la metanogenesi. La strategia utilizzata per convalidare il modello comunitario è stata valutare il tasso di crescita di ogni membro, in risposta all'aumento dei limiti di assorbimento di CO<sub>2</sub> e H<sub>2</sub>. Aumentando quindi la disponibilità di nutrienti/composti, è stato possibile evidenziare la dipendenza diretta delle specie per la CO<sub>2</sub>/H<sub>2</sub>. Queste

intuizioni non solo arricchiscono la nostra comprensione delle dinamiche microbiche, ma aprono anche la strada allo sviluppo di strategie innovative per migliorare i processi biotecnologici, fornendo metodi per manipolare le interazioni microbiche in modo da aumentare l'efficienza produttiva. Attraverso questi approfondimenti, l'articolo dimostra come l'avanzamento dei modelli computazionali possa trasformare la nostra capacità di ottimizzare i processi industriali e biotecnologici, migliorando l'efficienza e la sostenibilità delle tecnologie per l'upgrading del biogas

3.3 “Diverse electron carriers drive syntrophic interactions in an enriched anaerobic acetate-oxidizing consortium” di McDaniel et. al 2023

L'ultima proposta esamina il ruolo delle interazioni sintrofiche nell'ossidazione dell'acetato in ambienti anaerobici, mettendo in luce come i modelli metabolici abbiano modificato la nostra comprensione di queste complesse dinamiche. Si sottolinea l'importanza dell'ossidazione sintrofica dell'acetato (SAO, syntrophic acetate oxidation) in sistemi di digestione anaerobica tra i batteri SAO (SAOB), operanti a temperature termofile e/o con elevate concentrazioni di ammoniaca. Sottolineando che il percorso SAO è cruciale per la stabilità del processo di digestione anaerobica, essa non è termodinamicamente fattibile sotto condizioni standard, e quindi richiede la presenza di un archaea metanogeno per mantenere basso i livelli di formiato e/o di idrogeno. (Schink, 1997) Attraverso un esperimento di arricchimento di lunga durata, l'articolo descrive come sia stato possibile isolare una comunità microbica altamente specializzata in grado di convertire l'acetato in metano. Questo studio ha utilizzato approcci metagenomici e meta proteo genomici per tracciare il flusso di carbonio attraverso i membri della comunità, consentendo un dettagliato modello della comunità a livello metabolico. Le tecniche di sondaggio con isotopi stabili (SIP) hanno permesso di collegare l'identità genomica alla funzione metabolica, rivelando che le differenze nell'utilizzo di donatori di elettroni come l'idrogeno e il formiato da parte di diverse specie di *Methanothermobacter* influenzano significativamente la struttura e l'adattabilità della comunità. Questo evidenzia come la flessibilità nel partizionamento degli elettroni sia un fattore determinante nella dinamica comunitaria, governata da un mutualismo guidato dalla termodinamica. Al momento dell'incubazione SIP inoltre, il sequenziamento metagenomico ha indicato come *Methanothermobacter\_1* fosse il genoma più abbondante (50%), seguito da *Methanothermobacter\_2* (25%) e DTU068\_1 (6%). Il lavoro dimostra quindi l'efficacia dei modelli metabolici per comprendere meccanismi complessi nelle comunità microbiche e sottolinea l'importanza di questi strumenti nel prevedere e ottimizzare le interazioni microbiche per

migliorare i processi di conversione del biogas. Questi modelli avanzati offrono una finestra unica sulle intricate reti metaboliche che regolano i cicli biogeochimici e il flusso di energia attraverso gli ecosistemi naturali e ingegnerizzati, aprendo nuove strade per interventi biotecnologici mirati

## 4. Limiti e sviluppi futuri dei Modelli Metabolici

### 4.1 Analisi critica dei limiti dei modelli metabolici attuali nella previsione del comportamento microbico reale

Ad oggi, la maggior parte delle ricostruzioni metaboliche sono generate sulla base di una combinazione di annotazione del genoma e di database, con forte affidamento sul trasferimento di annotazioni basati su omologia di sequenza con specie note per gli organismi meno studiati. Sono stati anche sviluppati metodi per aiutare ad automatizzare questo processo, ma le ricostruzioni risultanti richiedono ancora una cura manuale.

Sono apparsi diversi metodi automatizzati che facilitano il processo di ricostruzione. Alcuni sono usati per mappare i geni nel genoma alle reazioni che formano una rete metabolica preliminare, come Pathway Tools 116 per citarne uno (Karp PD et al. 2002), ed altri che sono usati per raffinare le reti riempiendo nelle reazioni mancanti, ad esempio GapFind / GapFill120 (Satish Kumar et al. 2007).

L'innovazione dei metodi ha portato a migliorare le bozze di ricostruzioni ricavate dalla mappatura genica alla reazione tramite database, in quanto possono correggere informazioni errate o mancanti da database metabolici e/o annotazioni genomiche. Poiché i metodi automatizzati sfruttano basi di dati metaboliche e di trasporto, con annotazioni del genoma, gli errori si propagheranno nelle reti ricostruite. Viene qui proposta una tabella di problemi comunemente incontrati durante la ricostruzione automatizzata di modelli metabolici, la quale consiste inoltre in una guida per l'uso di tali metodi al fine di consentire la loro ottimale applicazione.

**Tabella 1.** Riporta in uno schema tripartito: le problematiche con le rispettive descrizioni ed i propri metodi risolutivi (possibili), riguardo alla modellazione di reti metaboliche. La tabella inoltre si divide, sempre secondo la stessa metodologia, in “Problematiche associate alle annotazioni genomiche” e “Problematiche associate ai database”. Si viene dunque a delineare un quadro molto dettagliato sulle limitazioni del caso. (Karp et al. 2002)

<b>Problematiche associate alle annotazioni genomiche</b>		
<b>Problematica</b>	<b>Descrizione</b>	<b>Metodi Risolutivi (possibili)</b>
I database di riferimento non sono continuamente aggiornati con nuove informazioni	Man mano che si trovano nuovi geni, le annotazioni del genoma più vecchio non vengono aggiornate. Il risultato è geni annotati in modo errato	L'automazione delle annotazioni può essere usata per analizzare nuovamente le annotazioni più vecchie del genoma (Ramy K. Aziz et al. 2008)
Annotazioni incorrette	Le annotazioni errate possono essere geni mancanti (dovuti all'ordinamento o agli errori di individuazione del gene dell'algoritmo) o le annotazioni errate del gene.	L'analisi dei modelli, tramite l'utilizzo di algoritmi, può contribuire a identificare alcuni di questi errori (Reed JL, et al. 2006) (Feist AM, et al. 2006)
Funzioni mancanti	Circa il 30% degli enzimi associati con numeri EC (Enzyme Commission) non presenta dati riguardo la sequenza dei geni che li codificano (Pouliot Y et al. 2007). Pertanto, non tutte le reazioni avranno sequenze di geni o proteine ad esse associate	Sono stati sviluppati strumenti automatizzati per individuare le reazioni mancanti
Specificità dei trasportatori	Le annotazioni per i trasportatori sono spesso carenti di dettagli sufficienti per determinare quale substrato(i) trasportano, anche se il meccanismo di funzionamento è noto	Sono necessarie metodologie per migliorare le annotazioni funzionali dei trasportatori
<b>Problematiche associate ai database</b>		
Associazioni gene-proteina-reazione	Le relazioni tra geni, enzimi e reazioni non sono sempre	Può essere automatizzato sulla base di confronti di

(GPR)	chiaramente definite (ad esempio, le sottounità di isoenzimi)	sequenze e delle GPRs conosciute (Notebaart RA et al. 2006)
Specificità delle reazioni	<p>Durante la mappatura delle reti metaboliche, sono due le cose da tenere conto,</p> <ol style="list-style-type: none"> <li>1) Specificità: La mancanza di specificità può condurre a confusione circa le vie e le interazioni metaboliche precise.</li> <li>2) Chiarezza del percorso: Definire chiaramente i composti specifici assicura la rappresentazione accurata dei percorsi metabolici e degli aiuti nella comprensione dei trattamenti biochimici esatti</li> </ol>	Sono necessari modifiche dei database .
Squilibri di reazione	La stechiometria di alcune reazioni non è bilanciata per quanto riguarda atomi di H, C, P, N, O o S.	Sono disponibili procedure automatizzate per controllare il bilanciamento delle reazioni (Pharkya P et al. 2004)
Stati di protonazione dei composti	Le reazioni sono generalmente scritte per la versione non protonata delle molecole. Questo influenza i coefficienti stechiometrici per i protoni nei modelli.	Software per la previsione del pKa sono strumenti che utilizzano modelli computazionali. Questi software sfruttano tecniche di ML, simulazioni quantistiche e banche dati sperimentali per prevedere con precisione i valori di pKa
Disponibilità di coenzimi	Gli enzimi hanno spesso bisogno di coenzimi. Affinché gli enzimi siano funzionali, la cellula deve essere in grado di produrli o ottenerli dall'ambiente.	Dati riguardo il ruolo di coenzimi in molteplici reazioni sono disponibili. Ad esempio, BRENDA ( Schomburg I. et al. 2004) contiene dati che possono essere sfruttati per l'automazione.

4.2 Discussione su possibili avanzamenti nella modellazione metabolica, inclusa l'integrazione di dati biologici più dinamici e approcci di apprendimento automatico.

C'è un crescente interesse nel combinare tecniche di ML con i GEMS. Gli approcci ML possono affinare i vincoli di input per i GEMS e confrontare i risultati delle simulazioni utilizzando dati sperimentali. Altre metodologie, tra cui la regressione lineare (LR), gli alberi decisionali e gli ingenui Bayes, sono stati utilizzati per colmare le lacune nei GEMS draft. Inoltre, possono essere integrati, come detto in precedenza, i metodi di modellazione basati sulla costrizione come FBA, incorporando ulteriori vincoli derivanti da dati omici, come i profili di espressione genica, l'abbondanza di proteine e le concentrazioni di metaboliti. Sfortunatamente, la quantificazione completa dei metaboliti nei campioni biologici non si traduce immediatamente nella conoscenza dell'utilizzo delle vie metaboliche. Stimando i flussi attraverso le vie metaboliche, i modelli computazionali mantengono la promessa di colmare questo divario tra i dati e la funzionalità biologica (Damiani et al. 2019). Uno strumento fondamentale per l'ispezione, l'interpretazione e lo sfruttamento dei dati omici e di conseguenza il progressivo miglioramento nella modellazione metabolica, è l'apprendimento automatico (dettagliato), che ha probabilmente contribuito a compiere diversi passi in avanti nelle recenti ricerche e si prevede che le guiderà sempre più nel prossimo futuro. L'apprendimento automatico può essere descritto come un insieme di algoritmi che migliorano l'accuratezza della previsione attraverso l'esperienza, dato un certo input processabile, da cui sono in grado di apprendere e generalizzare. (Zampieri et al. 2019) Parallelamente, l'aumento dei dati e delle conoscenze ha anche favorito lo sviluppo di modelli matematici per sistemi biomolecolari. Contrariamente agli approcci guidati dall'estrapolazione di dati, l'analisi ipotesi-guidata dei domini omici su grande scala rimane tipicamente proibitiva data la difficoltà nell'individuazione dei meccanismi biologici di fondo. Esistono tuttavia alcune eccezioni. Tra i vari approcci, la modellazione basata su vincoli (CBM, constraint based modeling) del metabolismo sta ricevendo un enorme impulso grazie alla sua ampia portata e flessibilità, consentendo intuizioni meccanicistiche nella relazione genotipo-fenotipo-ambiente, attraverso l'integrazione con dati omici.

#### 4.3 Come questi miglioramenti potrebbero portare a una migliore accuratezza nella previsione delle interazioni microbiche e dei comportamenti dei sistemi

Ad oggi, il campo dei GEMS testimonia la potenzialità ed il progresso che la modellazione metabolica può offrire attraverso l'applicazione ai vari domini principali, grazie all'abbondanza di dati biologici, ai progressi nella strumentazione automatizzata di ricostruzione e all'introduzione di nuove tecniche di modellazione matematica. Mentre i GEMS continuano ad evolversi, verrà compresa una gamma più ampia delle vie biologiche e delle associazioni gene-proteina-reazione (GPR). Per migliorare ulteriormente le loro capacità, è necessario incorporare ulteriori informazioni biochimiche nei GEMS, andando oltre il metabolismo. Ciò include funzioni quali la composizione macromolecolare cellulare, le informazioni strutturali dettagliate sulle proteine e la loro allocazione. Considerevolmente, determinati trattamenti molecolari quali le interazioni del enzima-substrato, le strutture complesse della proteina-proteina e le modifiche post-traduzionali richiedono un'attenta considerazione. Gli sforzi dovrebbero essere diretti verso la standardizzazione e l'armonizzazione dei dati omici, lo sviluppo di pipeline, strumenti automatizzati per l'integrazione, la normalizzazione dei dati e l'integrazione di algoritmi ML per migliorare la precisione e la robustezza. Gli algoritmi ML in particolare, possono svolgere un ruolo cruciale nell'automatizzare la ricostruzione dei GEMS, sfruttando i dati omici e le ricostruzioni di reti metaboliche esistenti. Inoltre, possono essere utilizzati per validare e perfezionare i GEMS, utilizzando dati sperimentali e garantendo una rappresentazione accurata delle caratteristiche metaboliche. (Sen e Oresic, 2023)

## 5. Conclusione

### 5.1 Riassunto delle principali scoperte della tesi

Nel corso di tutto l'elaborato, ma grazie in modo particolare alle analisi approfondite dei tre casi di studio proposti, abbiamo posto sotto lente d'ingrandimento i rapporti che si vengono a creare tra gli organismi in comunità batteriche complesse. Un argomento che sta particolarmente a cuore alla comunità scientifica, che potrebbe condurci sulla retta via nel processo di riduzione dell'utilizzo dei combustibili fossili, se pensiamo in un'ottica terrestre. Ma che allargando gli orizzonti, può aiutare anche nella ricerca della vita sul pianeta Marte, che sarà probabilmente facilitata da una comprensione degli ecosistemi sotterranei sulla Terra che meglio simulano quelli sul pianeta rosso. La

misura in cui i protagonisti di questa tesi, gli archaea metanogeni, contribuiscono alla produzione annuale di CH<sub>4</sub> sulla Terra non è nota, anche se potenzialmente significativa.

Gli ambienti del sottosuolo metanogenico della Terra sono la migliore approssimazione degli ambienti su Marte che più probabilmente ospitano metanogeni; quindi, una comprensione biochimica di queste vie dovrebbe fornire una base per la progettazione di esperimenti per rilevare forme di vita su Marte produttori di metano (Ferry 2010).

In modo più dettagliato, viene evidenziata l'importanza delle relazioni simbiotiche nella produzione finale di gas biocompatibili e dei modelli metabolici nella digestione anaerobica.

Si prevede infatti che il processo di ricostruzione delle reti metaboliche continuerà a crescere in portata, profondità e precisione, e dovrebbe continuare a consentire un ampliamento dello spettro di studi di base e applicati. La disponibilità di ricostruzioni complete di alta qualità accelererà l'implementazione del paradigma di biologia dei sistemi (cioè, componenti biologici a reti a modelli computazionali a studi fenotipici) e contribuirà così a realizzare l'ampio potenziale di trasformazione di questo teorema nelle scienze della vita.

Abbiamo a tal proposito, proposto nella nostra analisi un numero considerevole di approcci computazionali all'avanguardia, indicati per snellire il processo di studio riguardo ai contributi metabolici degli operatori microbici, all'interno di una nicchia ecologica e all'estensione della direzionalità delle interazioni tra di loro. Entrando nel dettaglio, queste descrizioni multi-livello permettono di quantificare correttamente i compromessi tra le forze motrici egoiste e altruistiche in un ecosistema microbico, tuttavia, notiamo che la fisiologia delle comunità microbiche è altamente dipendente dal contesto e dall'ambiente e un criterio di idoneità universale specifico della comunità non esiste. Difatti è stato inevitabile precisare che i sistemi biologici funzionino attraverso interazioni complesse fra varie omiche ed una comprensione più completa di questi sistemi sia possibile soltanto assumendo una prospettiva multi-omica mirata. Tutto ciò, ha presentato di conseguenza la necessità per lo sviluppo di approcci di integrazione, che sono in grado di catturare la complessità delle interazioni che definiscono questi sistemi biologici e sono adattate alle sfide di combinazione dei dati eterogenei attraverso punti di vista differenti.

A causa dei costi non accessibili a tutti, alla sensibilità degli strumenti o ad altri fattori sperimentali, i dati per un campione biologico possono mancare per una o più "omic technologies". A supporto di ciò, i recenti sviluppi metodologici nell'intelligenza artificiale (AI) e nell'apprendimento statistico hanno notevolmente facilitato le analisi di dati omici. Tuttavia molte di queste tecniche

richiedono l'accesso a dati studiabili nel dettaglio e nell'applicare questi metodi, gli utenti dovrebbero essere consapevoli dei loro limiti.

Precisiamo “dati studiabili nel dettaglio”, in quanto gli ostacoli ai metodi di integrazione efficaci includono, ma non sono limitati a, la natura eterogenea dei dati e delle proprietà distributive attraverso i set di dati multi-omici, la differenza tipicamente grande nel numero delle biomolecole misurate confrontate al numero dei campioni o delle repliche, e la natura complessa spesso rumorosa dei dati biologici.

5.2 Riflessione sull'importanza dei modelli metabolici nel progredire la nostra comprensione delle interazioni sintrofiche e le loro implicazioni più ampie nella biotecnologia.

I metanogeni sono biocatalizzatori, ovvero che hanno il potenziale per contribuire a una soluzione per problemi energetici futuri, producendo metano come vettore di energia immagazzinabile. Questo può essere utilizzato come vettore di energia stoccabile, come carburante per veicoli, per la produzione di energia elettrica, o come sostanza chimica di base per la sintesi e molti paesi hanno già reti di gas naturale ben sviluppate. Negli impianti di biogas, sono state intraprese misurazioni per aumentare la resa e la purezza del biogas, come l'aggiunta di idrogeno o granulato metallico. Queste strutture sono tipicamente composte da diverse sezioni, che possono includere vasche o serbatoi, ciascuno destinato a ospitare il materiale in fase di digestione. La progettazione di questi impianti prevede una suddivisione strategica delle diverse fasi della digestione anaerobica, in modo da assicurare che i parametri essenziali, come temperatura, pH e tempi di ritenzione, siano mantenuti a livelli ottimali per il progresso delle reazioni biologiche. Questa configurazione modulare permette di gestire e ottimizzare ogni fase del processo, dal pretrattamento all'idrolisi, passando per l'acidogenesi e l'acetogenesi, fino alla metanogenesi finale. In questo modo, si garantisce un ambiente controllato che favorisce l'efficienza e la stabilità del sistema, migliorando la produzione di biogas e riducendo i rischi di interferenze o fallimenti operativi. La metanazione biologica si trova naturalmente nelle paludi, nei sistemi digestivi degli animali, nei giacimenti petroliferi e in altri ambienti ed è già comunemente utilizzata negli impianti di acque reflue e negli impianti di biogas. Le nuove applicazioni per i metanogeni come l'elettrometanogenesi sono in aumento, ma tuttavia, c'è ancora un sacco di ricerca di base in corso sul gruppo molto diversificato e unico dei metanogeni.

### 5.3 Utilizzo ed applicazione dei metanogeni

Gli Archaea metanogeni rappresentano un gruppo estremamente variegato, con ceppi che possono prosperare in condizioni ambientali molto difficili, come temperature estremamente alte o basse, alti livelli di osmolarità e valori di pH estremi. Di conseguenza, è cruciale sviluppare e ottimizzare i processi industriali che sfruttano questi microrganismi. Questa ottimizzazione non è importante solo per la produzione di CH<sub>4</sub>, ma anche per la produzione di altri prodotti e per varie applicazioni industriali, qui di seguito verranno riportate le più importanti :

#### 1. Produzione di Idrogeno,

È stato osservato che diversi ceppi metanogenici possono anche produrre idrogeno (Valentine et al. 2000; Goyal et al. 2016). Questo può accadere se la quantità di idrogeno disponibile è limitata (sub-nanomolare); si è scoperto che non il metano, ma il formato e possibilmente altri metaboliti possono essere la fonte di H<sub>2</sub>. Questa applicazione è ancora limitata alla scala di laboratorio e per creare un processo ragionevole, l'ingegneria genetica dovrebbe essere fatta per aumentare la resa di idrogeno. Si presume che le idrogenasi presenti nei metanogeni siano gli enzimi che catalizzano la produzione di idrogeno ed un modo possibile per aumentare la resa di idrogeno potrebbe quindi essere l'individuazione della relativa idrogenasi e successivamente sovraesprimerla.

#### 2. Produzione biotecnologica mediante metanogeni geneticamente modificati

Negli ultimi anni, la strumentazione genetica per i metanogeni è stata migliorata, aprendo un nuovo campo di ricerca su questi importanti microrganismi. Come primo passo, lo spettro del prodotto di metanogeni potrebbe essere aumentato. Oltre a consentire prodotti diversi, è stato anche possibile ampliare la gamma di substrati. Tuttavia, poiché sia i tassi di conversione che i rendimenti dei prodotti erano bassi e in nessun caso le stechiometrie di conversione riportate, l'applicabilità di tale sistema rimane in discussione. Lo stesso vale per la produzione di altri prodotti di alto valore come aminoacidi o vitamine con metanogeni, che a causa della loro lenta crescita, non è ancora stata sviluppata un'applicazione tecnica (Schiraldi et al. 2002). Ma poiché c'è un continuo progresso nello sviluppo di strumenti genetici per metanogeni, come descritto sopra, è pensabile che nuovi processi con metanogeni eterologhi emergeranno nei prossimi anni.

#### 3. Metano dai giacimenti di petrolio e carbone

Quasi due terzi del petrolio fossile rimane all'interno dei campi petroliferi se si utilizzano metodi di produzione convenzionali (Gieg et al. 2008). È

stato osservato che il petrolio residuo può essere convertito in gas naturale da un consorzio metanogenico, che è stato aggiunto al campo petrolifero (Gieg et al. 2008). Oltre ai giacimenti petroliferi, anche le sabbie bituminose o altre emulsioni olio-acqua potrebbero essere trattate in questo modo (Voordouw 2011). Ma poiché i costi per il gas naturale rimangono relativamente bassi, mentre quelli per il petrolio greggio sono significativamente più elevati, questo approccio rimane sperimentale a causa della mancanza di benefici (Voordouw 2011). Una fonte naturale di metano è il metano del letto di carbone. È stato scoperto che circa il 40% di questo metano è prodotto da consorzi microbici contenenti metanogeni (ad es. Methanosarcinales).

#### 4. Produzione di biogas da materia organica

L'ultima ma principale applicazione tecnica dei metanogeni è la produzione di biogas mediante digestione di substrati organici. A gran parte, già ripresa nel corso di questo elaborato, la digestione è vista come un processo a quattro fasi.

## Bibliografia

- 1) Abdelsalam, E. *et al.* (2017) 'Influence of zero valent iron nanoparticles and magnetic iron oxide nanoparticles on biogas and methane production from anaerobic digestion of manure', *Energy*, 120, pp. 842–853. Available at: <https://doi.org/10.1016/j.energy.2016.11.137>.
- 2) Anderson, J.K., Smith, T.G. and Hoover, T.R. (2010) 'Sense and sensibility: flagellum-mediated gene regulation', *Trends in Microbiology*, 18(1), pp. 30–37. Available at: <https://doi.org/10.1016/j.tim.2009.11.001>.
- 3) Aziz, R.K. *et al.* (2008) 'The RAST Server: Rapid Annotations using Subsystems Technology', *BMC Genomics*, 9(1), p. 75. Available at: <https://doi.org/10.1186/1471-2164-9-75>.

- 4) Babu, M. and Snyder, M. (2023) 'Multi-Omics Profiling for Health', *Molecular & Cellular Proteomics : MCP*, 22(6), p. 100561. Available at: <https://doi.org/10.1016/j.mcpro.2023.100561>.
- 5) Blazier, A.S. and Papin, J.A. (2012) 'Integration of expression data in genome-scale metabolic network reconstructions', *Frontiers in Physiology*, 3, p. 299. Available at: <https://doi.org/10.3389/fphys.2012.00299>.
- 6) Borodina, I. and Nielsen, J. (2005) 'From genomes to in silico cells via metabolic networks', *Current Opinion in Biotechnology*, 16(3), pp. 350–355. Available at: <https://doi.org/10.1016/j.copbio.2005.04.008>.
- 7) Boucher, D.H. (1985) *The Biology of Mutualism: Ecology and Evolution*. Oxford University Press.
- 8) Breitling, R., Vitkup, D. and Barrett, M.P. (2008a) 'New surveyor tools for charting microbial metabolic maps', *Nature Reviews Microbiology*, 6(2), pp. 156–161. Available at: <https://doi.org/10.1038/nrmicro1797>.
- 9) Breitling, R., Vitkup, D. and Barrett, M.P. (2008b) 'New surveyor tools for charting microbial metabolic maps', *Nature Reviews Microbiology*, 6(2), pp. 156–161. Available at: <https://doi.org/10.1038/nrmicro1797>.
- 10) Brunk, E. *et al.* (2018) 'Recon3D enables a three-dimensional view of gene variation in human metabolism', *Nature Biotechnology*, 36(3), pp. 272–281. Available at: <https://doi.org/10.1038/nbt.4072>.
- 11) Damiani, C. *et al.* (2019) 'Integration of single-cell RNA-seq data into population models to characterize cancer metabolism', *PLOS Computational Biology*, 15(2), p. e1006733. Available at: <https://doi.org/10.1371/journal.pcbi.1006733>.
- 12) Dinopoulou, G., Rudd, T. and Lester, J.N. (1988) 'Anaerobic acidogenesis of a complex wastewater: I. The influence of operational parameters on reactor performance', *Biotechnology and Bioengineering*, 31(9), pp. 958–968. Available at: <https://doi.org/10.1002/bit.260310908>.
- 13) Enzmann, F. *et al.* (2018) 'Methanogens: biochemical background and biotechnological applications', *AMB Express*, 8(1), p. 1. Available at: <https://doi.org/10.1186/s13568-017-0531-x>.

- 14) Feist, A.M. *et al.* (2006) 'Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*', *Molecular Systems Biology*, 2, p. 2006.0004. Available at: <https://doi.org/10.1038/msb4100046>.
- 15) Feist, A.M. *et al.* (2009) 'Reconstruction of Biochemical Networks in Microbial Organisms', *Nature reviews. Microbiology*, 7(2), pp. 129–143. Available at: <https://doi.org/10.1038/nrmicro1949>.
- 16) Ferry, J.G. (2010) 'The chemical biology of methanogenesis', *Planetary and Space Science*, 58(14), pp. 1775–1783. Available at: <https://doi.org/10.1016/j.pss.2010.08.014>.
- 17) Flores, J.E. *et al.* (2023) 'Missing data in multi-omics integration: Recent advances through artificial intelligence', *Frontiers in Artificial Intelligence*, 6. Available at: <https://doi.org/10.3389/frai.2023.1098308>.
- 18) *Food Manufacturers* (no date) *Denali Water Solutions*. Available at: <https://www.denalicorp.com/markets/food-manufacturers/> (Accessed: 30 May 2024).
- 19) Fournier, G.P. and Gogarten, J.P. (2008) 'Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*', *Journal of Bacteriology*, 190(3), pp. 1124–1127. Available at: <https://doi.org/10.1128/JB.01382-07>.
- 20) Gieg, L.M., Duncan, K.E. and Suflita, J.M. (2008) 'Bioenergy production via microbial conversion of residual oil to natural gas', *Applied and Environmental Microbiology*, 74(10), pp. 3022–3029. Available at: <https://doi.org/10.1128/AEM.00119-08>.
- 21) Gomez-Cabrero, D. *et al.* (2014) 'Data integration in the era of omics: current and future challenges', *BMC Systems Biology*, 8(2), p. 11. Available at: <https://doi.org/10.1186/1752-0509-8-S2-I1>.
- 22) Goyal, N., Zhou, Z. and Karimi, I.A. (2016) 'Metabolic processes of *Methanococcus maripaludis* and potential applications', *Microbial Cell Factories*, 15(1), p. 107. Available at: <https://doi.org/10.1186/s12934-016-0500-0>.

- 23) Hamid: *Data integration in genetics and genomics:...* - Google Scholar (no date). Available at: <https://scholar.google.com/scholar> (Accessed: 31 May 2024).
- 24) Hamid, J.S. *et al.* (2009) 'Data Integration in Genetics and Genomics: Methods and Challenges', *Human Genomics and Proteomics : HGP*, 2009, p. 869093. Available at: <https://doi.org/10.4061/2009/869093>.
- 25) *Integrating metagenomic binning with flux balance analysis to unravel syntrophies in anaerobic CO2 methanation | Microbiome | Full Text* (no date). Available at: <https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01311-1> (Accessed: 16 May 2024).
- 26) Jackson, B.E. and McInerney, M.J. (2002) 'Anaerobic microbial metabolism can proceed close to thermodynamic limits', *Nature*, 415(6870), pp. 454–456. Available at: <https://doi.org/10.1038/415454a>.
- 27) Kanehisa, M. *et al.* (2008) 'KEGG for linking genomes to life and the environment', *Nucleic Acids Research*, 36(Database issue), pp. D480-484. Available at: <https://doi.org/10.1093/nar/gkm882>.
- 28) Karp, P.D., Paley, S. and Romero, P. (2002) 'The Pathway Tools software', *Bioinformatics (Oxford, England)*, 18 Suppl 1, pp. S225-232. Available at: [https://doi.org/10.1093/bioinformatics/18.suppl\\_1.s225](https://doi.org/10.1093/bioinformatics/18.suppl_1.s225).
- 29) Kouzuma, A., Kato, S. and Watanabe, K. (2015) 'Microbial interspecies interactions: recent findings in syntrophic consortia', *Frontiers in Microbiology*, 6. Available at: <https://doi.org/10.3389/fmicb.2015.00477>.
- 30) Krassowski, M. *et al.* (2020) 'State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing', *Frontiers in Genetics*, 11. Available at: <https://doi.org/10.3389/fgene.2020.610798>.
- 31) Lin, L. *et al.* (2010) 'In-depth investigation of enzymatic hydrolysis of biomass wastes based on three major components: Cellulose, hemicellulose and lignin', *Bioresource Technology*, 101(21), pp. 8217–8223. Available at: <https://doi.org/10.1016/j.biortech.2010.05.084>.

- 32) Machado, D. *et al.* (2018) 'Fast automated reconstruction of genome-scale metabolic models for microbial species and communities', *Nucleic Acids Research*, 46(15), pp. 7542–7553. Available at: <https://doi.org/10.1093/nar/gky537>.
- 33) McDaniel, E.A. *et al.* (2023) 'Diverse electron carriers drive syntrophic interactions in an enriched anaerobic acetate-oxidizing consortium', *The ISME Journal*, 17(12), pp. 2326–2339. Available at: <https://doi.org/10.1038/s41396-023-01542-6>.
- 34) Meegoda, J.N. *et al.* (2018) 'A Review of the Processes, Parameters, and Optimization of Anaerobic Digestion', *International Journal of Environmental Research and Public Health*, 15(10), p. 2224. Available at: <https://doi.org/10.3390/ijerph15102224>.
- 35) Meyer, F. *et al.* (2008) 'The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes', *BMC bioinformatics*, 9, p. 386. Available at: <https://doi.org/10.1186/1471-2105-9-386>.
- 36) Notebaart, R.A. *et al.* (2006) 'Accelerating the reconstruction of genome-scale metabolic networks', *BMC Bioinformatics*, 7, p. 296. Available at: <https://doi.org/10.1186/1471-2105-7-296>.
- 37) Orth, J.D., Thiele, I. and Palsson, B.Ø. (2010) 'What is flux balance analysis?', *Nature Biotechnology*, 28(3), pp. 245–248. Available at: <https://doi.org/10.1038/nbt.1614>.
- 38) Pharkya, P., Burgard, A.P. and Maranas, C.D. (2004) 'OptStrain: a computational framework for redesign of microbial production systems', *Genome Research*, 14(11), pp. 2367–2376. Available at: <https://doi.org/10.1101/gr.2872004>.
- 39) Pharkya, P. and Maranas, C.D. (2006) 'An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems', *Metabolic Engineering*, 8(1), pp. 1–13. Available at: <https://doi.org/10.1016/j.ymben.2005.08.003>.
- 40) Picard, M. *et al.* (2021) 'Integration strategies of multi-omics data for machine learning analysis', *Computational and Structural Biotechnology Journal*, 19, pp. 3735–3746. Available at: <https://doi.org/10.1016/j.csbj.2021.06.030>.

- 41) Pouliot, Y. and Karp, P.D. (2007) 'A survey of orphan enzyme activities', *BMC bioinformatics*, 8, p. 244. Available at: <https://doi.org/10.1186/1471-2105-8-244>.
- 42) Price, N.D., Reed, J.L. and Palsson, B.Ø. (2004) 'Genome-scale models of microbial cells: evaluating the consequences of constraints', *Nature Reviews Microbiology*, 2(11), pp. 886–897. Available at: <https://doi.org/10.1038/nrmicro1023>.
- 43) Reed, J.L., Patel, T.R., *et al.* (2006) 'Systems approach to refining genome annotation', *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), pp. 17480–17484. Available at: <https://doi.org/10.1073/pnas.0603364103>.
- 44) Reed, J.L., Famili, I., *et al.* (2006) 'Towards multidimensional genome annotation', *Nature Reviews. Genetics*, 7(2), pp. 130–141. Available at: <https://doi.org/10.1038/nrg1769>.
- 45) Reed, J.L. and Palsson, B.Ø. (2003) 'Thirteen Years of Building Constraint-Based In Silico Models of Escherichia coli', *Journal of Bacteriology* [Preprint]. Available at: <https://doi.org/10.1128/jb.185.9.2692-2699.2003>.
- 46) Satish Kumar, V., Dasika, M.S. and Maranas, C.D. (2007) 'Optimization based automated curation of metabolic reconstructions', *BMC Bioinformatics*, 8(1), p. 212. Available at: <https://doi.org/10.1186/1471-2105-8-212>.
- 47) Schink, B. (1997) 'Energetics of syntrophic cooperation in methanogenic degradation', *Microbiology and molecular biology reviews: MMBR*, 61(2), pp. 262–280. Available at: <https://doi.org/10.1128/membr.61.2.262-280.1997>.
- 48) Schiraldi, C., Giuliano, M. and De Rosa, M. (2002) 'Perspectives on biotechnological applications of archaea', *Archaea (Vancouver, B.C.)*, 1(2), pp. 75–86. Available at: <https://doi.org/10.1155/2002/436561>.
- 49) Schomburg, I. *et al.* (2004) 'BRENDA, the enzyme database: updates and major new developments', *Nucleic Acids Research*, 32(Database issue), pp. D431-433. Available at: <https://doi.org/10.1093/nar/gkh081>.
- 50) Sen, P. and Orešič, M. (2023a) 'Integrating Omics Data in Genome-Scale Metabolic Modeling: A Methodological Perspective for

Precision Medicine', *Metabolites*, 13(7), p. 855. Available at:  
<https://doi.org/10.3390/metabo13070855>.

- 51) Sen, P. and Orešič, M. (2023b) 'Integrating Omics Data in Genome-Scale Metabolic Modeling: A Methodological Perspective for Precision Medicine', *Metabolites*, 13(7), p. 855. Available at:  
<https://doi.org/10.3390/metabo13070855>.
- 52) Stolyar, S. *et al.* (2007) 'Metabolic modeling of a mutualistic microbial community', *Molecular Systems Biology*, 3(1), p. 92. Available at:  
<https://doi.org/10.1038/msb4100131>.
- 53) *Sustainability | Free Full-Text | Comparative Analysis of Urban Development Trends of Beijing and Karachi Metropolitan Areas* (no date). Available at: <https://www.mdpi.com/2071-1050/12/2/451> (Accessed: 30 May 2024).
- 54) Thiele, I. and Palsson, B.Ø. (2010) 'A protocol for generating a high-quality genome-scale metabolic reconstruction', *Nature Protocols*, 5(1), pp. 93–121. Available at:  
<https://doi.org/10.1038/nprot.2009.203>.
- 55) Valentine, D.L., Blanton, D.C. and Reeburgh, W.S. (2000) 'Hydrogen production by methanogens under low-hydrogen conditions', *Archives of Microbiology*, 174(6), pp. 415–421. Available at:  
<https://doi.org/10.1007/s002030000224>.
- 56) Wolfe, R.S. (2011) 'Techniques for cultivating methanogens', *Methods in Enzymology*, 494, pp. 1–22. Available at:  
<https://doi.org/10.1016/B978-0-12-385112-3.00001-9>.
- 57) Zampieri, G. *et al.* (2019) 'Machine and deep learning meet genome-scale metabolic modeling', *PLOS Computational Biology*, 15(7), p. e1007084. Available at:  
<https://doi.org/10.1371/journal.pcbi.1007084>.
- 58) Zarayeneh, N. *et al.* (2017) 'Integration of multi-omics data for integrative gene regulatory network inference', *International Journal of Data Mining and Bioinformatics*, 18(3), p. 223. Available at:  
<https://doi.org/10.1504/IJDMB.2017.087178>.
- 59) Zhu, X. *et al.* (2020) 'Metabolic dependencies govern microbial syntrophies during methanogenesis in an anaerobic digestion

ecosystem', *Microbiome*, 8, p. 22. Available at:  
<https://doi.org/10.1186/s40168-019-0780-9>.

- 60) Zomorodi, A.R. and Maranas, C.D. (2012) 'OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities', *PLoS computational biology*, 8(2), p. e1002363. Available at:  
<https://doi.org/10.1371/journal.pcbi.1002363>.